



1.1 ความเป็นมาของปัญหา

ปัจจุบันเครื่องคอมพิวเตอร์ถูกนำมาใช้กันอย่างแพร่หลาย เพื่อช่วยให้การประมวลผลให้เป็นไปอย่างรวดเร็ว และถูกต้อง โดยปกติ การติดต่อระหว่างเครื่องคอมพิวเตอร์กับมนุษย์จะทำโดย เครื่องจะรับคำสั่งหรือข้อมูลทางแป้นพิมพ์ และจะแสดงผลทางจอภาพ หรือเครื่องพิมพ์ การติดต่อกับเครื่องโดยผ่านทางแป้นพิมพ์ต้องอาศัยความชำนาญ และการฝึกฝนเป็นเวลานาน จึงจะสามารถใช้เครื่องคอมพิวเตอร์ได้อย่างมีประสิทธิภาพ

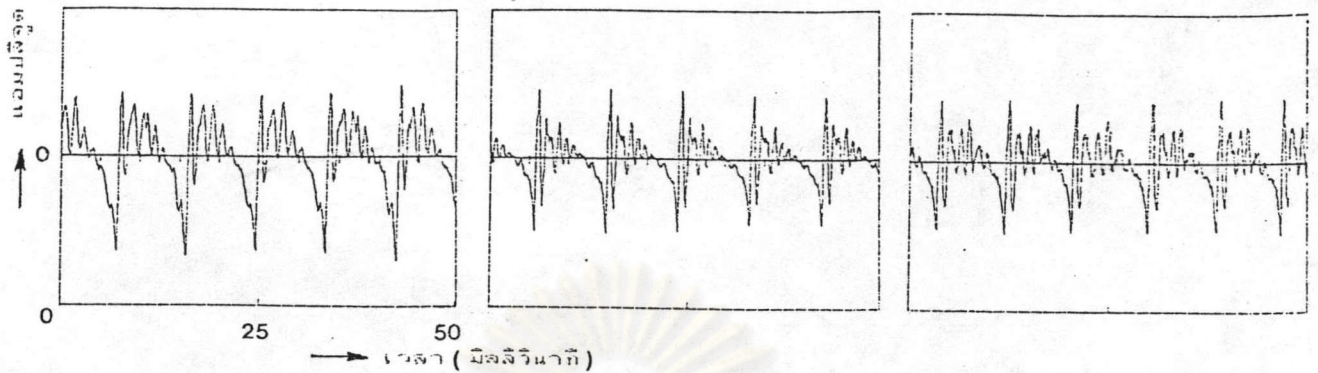
เนื่องจากมนุษย์มีความต้องการที่จะเพิ่มความสามารถในการติดต่อกับเครื่อง (Man-machine Communication) และ โดยปกติ มนุษย์มักจะสื่อสารกันด้วยเสียงพูด ทั้งนี้เพราะเป็นวิธีที่สะดวก รวดเร็ว และปัญหาที่เกิดขึ้นจากการใช้แป้นพิมพ์ ทำให้การติดต่อกับเครื่องโดยใช้เสียงพูดเป็นที่ต้องการประการหนึ่งของคนเรา

ปัญหาพื้นฐาน 2 ประการสำหรับการรับรู้เสียงพูดด้วยคอมพิวเตอร์ คือ ประการแรก ในแต่ละคำพูดของคนเราที่เปล่งออกมา จะมีลักษณะเฉพาะตัวเอง แม้ว่าเราจะพยายามเปล่งคำพูดให้มีเสียงใกล้เคียงกันเท่าใดก็ตาม จากรูปที่ 1.1 จะแสดงให้เห็นถึง รูปคลื่นของเสียง "เอ" ที่พูดจากบุคคลเดียวกันหลาย ๆ ครั้ง เราจะพบถึงความแตกต่างที่เกิดขึ้นได้อย่างชัดเจน (Bunge E., 1977)

ประการที่สอง เนื่องจากเสียงของคนเราที่ใช้ในการสื่อความหมายต่าง ๆ นั้น จะอยู่ในช่วงความถี่ 4-5 กิโลเฮิร์ตซ์ และจากทฤษฎีการสุ่มตัวอย่าง (Sampling Theorem) พบว่า ในการที่จะรักษาความหมายของเสียงที่พูดออกมา 1 วินาที ให้สมบูรณ์นั้น จะต้องใช้อัตราการสุ่มตัวอย่าง อย่างน้อย 2 เท่าของความถี่เสียง (Carlson A.B., 1975) ซึ่งจะเท่ากับ 8,000-10,000 ค่าเป็นอย่างน้อย ในการรับรู้เสียงพูดของเครื่องนั้นจำเป็นต้องใช้เสียงพูดเป็นเวลาหลายวินาทีและผู้พูดหลายคน ดังนั้นจึงมีข้อมูลจำนวนมากที่จะต้องถูกเก็บและนำไปใช้ ซึ่งจะมีผลต่อสื่อที่ใช้ในการบันทึก และเวลาที่ใช้ในการรับรู้ เทคนิคของการวิเคราะห์ให้ได้พารามิเตอร์ของสัญญาณ เพื่อลดจำนวนข้อมูลลงจึงเป็นทางออกประการหนึ่งในการแก้ปัญหา

เราสามารถแบ่งวิธีการรับรู้เสียงพูดออกได้เป็น 2 วิธี คือ (วีระ รวีพิทักษ์, 2529)

ก. แบบพิจารณาทั้งหน่วยภาษาที่เปล่งออกมาทั้งหมด (Isolated Words Recognition) หน่วยภาษาที่พิจารณาในวิธีนี้ อาจจะเป็นพยางค์เดี่ยว คำ กลุ่มคำ วลี หรือ ประโยค ก็ได้ วิธีนี้มีผลดี คือ จะหลีกเลี่ยงผลกระทบอันเกิดจากฐานของเสียงภายในคำ หรือกลุ่มคำนั้น



(ก)

(ข)

(ค)

รูปที่ 1.1 แสดงถึงรูปคลื่นเสียงของเสียง "เอ" ที่พูดจากบุคคลเดียวกัน 3 ครั้ง
(Bunge E., 1977)

ข. การรับรู้โดยแยกแยะรายละเอียดของหน่วยเสียง (Phonetic Recognition) ในวิธีนี้จะพิจารณาลักษณะของหน่วยเสียงที่มีขนาดเล็กลงไป เช่น ระดับสระ พยัญชนะ หรือวรรณยุกต์ เป็นต้น โดยจะใช้หน่วยเสียงย่อยเหล่านี้เป็นหลักในการเปรียบเทียบ วิธีนี้มีความสะดวกสำหรับระบบที่จะรับรู้จำนวนคำมาก ๆ การเปรียบเทียบข้อดี ข้อเสียของการเลือกหน่วยเสียง แสดงได้ ตารางที่ 1.1

จากวิธีการทั้งสองนี้ เราสามารถจำแนกวิธีการรับรู้เสียงพูดออกได้ตามลักษณะของคำพูดที่รับรู้ คือ

- ก. การรับรู้คำเดี่ยว (Isolated Words)
- ข. การรับรู้คำพูดต่อเนื่อง (Continuous Words)

และยังจำแนกได้ตามผู้พูด คือ

- ก. การรับรู้เสียงพูดเพียงคนเดียว (Speaker Dependent)
- ข. การรับรู้เสียงพูดแบบต่างบุคคล (Speaker Independent)

ซึ่งสามารถแสดงความสัมพันธ์ได้ ดังตารางที่ 1.2

ในการวิจัยนี้มุ่งเน้นที่การรับรู้กลุ่มคำเดี่ยวแบบต่างบุคคล ซึ่งจะมีปัญหาที่ก่อกำเนิดเพิ่มขึ้นอีกหลายประการ เช่น ผู้พูดแต่ละคนจะมีลักษณะเสียงเฉพาะของตนเอง ความเร็วในการออกเสียง การเน้นเสียง รวมทั้งสำเนียงที่แตกต่างกันอีกด้วย

ข้อดี	ข้อเสีย
<p><u>คำ</u></p> <ul style="list-style-type: none"> - ไม่มีผลกระทบอันเกิดจากฐานของเสียงภายใน - ใช้ได้กับทุกภาษา และง่าย <p><u>พยางค์</u></p> <ul style="list-style-type: none"> - ง่ายต่อการแยก - ไม่มีผลกระทบอันเกิดจากฐานของเสียงภายใน <p><u>หน่วยเสียง (phoneme)</u></p> <ul style="list-style-type: none"> - จำนวนคำเปรียบเทียบในหน่วยความจำน้อย - คำง่ายต่อการสร้างด้วยโฟเน็ม <p><u>หน่วยเสียงย่อย (allophone)</u></p> <ul style="list-style-type: none"> - ง่ายต่อการหาและตรวจพบ 	<ul style="list-style-type: none"> - จำนวนคำจำกัด - ยากต่อการปรับกับแต่ละคน <ul style="list-style-type: none"> - จำนวนหน่วยเสียงหลักที่จะเปรียบเทียบเพิ่มขึ้น - ยากต่อการหาขอบเขตของพยางค์ <ul style="list-style-type: none"> - ปัญหาการแยกชั้นกับความหมาย - จำนวนกฎเพื่อหาและแยกมีมาก <ul style="list-style-type: none"> - จำนวนหน่วยเสียงหลักที่จะเปรียบเทียบสูง

ตารางที่ 1.1 แสดงการเปรียบเทียบข้อดี ข้อเสียของการเลือกหน่วยภาษา
(วีระ รั้วพิทักษ์, 2529)

	ผู้พูดคนเดียว	แบบต่างบุคคล
วิธีการพิจารณาทั้งชุด	คำเดี่ยว	คำต่อเนื่อง
วิธีแยกแยะ	คำเดี่ยว	คำต่อเนื่อง

ตารางที่ 1.2 แสดงถึงวิธีการจำแนกวิธีการรับรู้เสียงพูดต่าง ๆ
(วีระ รั้วพิทักษ์, 2529)

1.2 วัตถุประสงค์ของวิทยานิพนธ์

1. เพื่อพัฒนาระบบให้สามารถรับรู้เสียงพูดที่เป็นกลุ่มคำเดี่ยวแบบต่างบุคคล
2. เพื่อหาอัลกอริทึมที่เหมาะสมสำหรับการเพิ่มอัตราความถูกต้องในการรับรู้
3. เพื่อประยุกต์ทฤษฎีต่าง ๆ เพื่อช่วยในการรับรู้ เช่น ไดนามิกโปรแกรมมิ่ง การสร้างแบบอ้างอิง เป็นต้น

1.3 ขอบเขตวิทยานิพนธ์

1. ทำการศึกษา และทดลองเพื่อหากรรมวิธีที่เหมาะสม และมีประสิทธิภาพในการรับรู้เสียงพูดกลุ่มคำโดดแบบต่างบุคคล

1.4 ขั้นตอนและวิธีดำเนินการ

1. ศึกษากระบวนการรับรู้เสียงพูดในลักษณะต่าง ๆ ที่ได้มีการศึกษามาแล้ว
2. คัดเลือกกรรมวิธีที่เหมาะสมในการรับรู้เสียงพูดกลุ่มคำเดี่ยวแบบต่างบุคคล โดยคำนึงถึงเวลาในการตอบสนอง และความถูกต้องในการรับรู้
3. คัดเลือกคำที่จะใช้ทดสอบในการรับรู้ โดยอาศัยหลักเกณฑ์ต่าง ๆ ที่เหมาะสม
4. พัฒนาอุปกรณ์เพื่อใช้ในการเปลี่ยนสัญญาณเสียงพูด ไปเป็นสัญญาณดิจิทัล เพื่อให้เครื่องคอมพิวเตอร์สามารถนำไปใช้งานได้
5. พัฒนาโปรแกรมสำหรับการวิเคราะห์สัญญาณเสียง การสร้างแบบอ้างอิง และการทดสอบการรับรู้
6. ทำการทดสอบ และปรับปรุงคุณภาพในการรับรู้ รวบรวมข้อมูล และประเมินผล

1.5 ประโยชน์ที่คาดว่าจะได้รับจากการทำวิทยานิพนธ์นี้

1. เป็นแนวทางในการพัฒนาระบบการรับรู้เสียงพูดในลักษณะอื่น ๆ ต่อไป เช่น การรับรู้เสียงพูดแบบคำต่อเนื่อง เป็นต้น
2. สามารถนำไปประยุกต์ใช้ในงานด้านต่าง ๆ เพื่อเพิ่มความสามารถในการติดต่อกับเครื่องโดยใช้เสียงพูด เช่น การนำไปเชื่อมต่อกับเวิร์ดโปรเซสเซอร์
3. ในส่วนของระบบประมวลผล TMS32010 ที่พัฒนาขึ้น สามารถนำไปประยุกต์ในงานด้านต่าง ๆ ได้เป็นอย่างดี โดยเฉพาะงานที่ต้องการความเร็วในการคำนวณสูง เช่น การวิเคราะห์สัญญาณเสียง เป็นต้น